UNIVERSITÉ DE RENNES 1

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2023

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : MathSTIC **POLE DE RECHERCHE :** Mathématiques-Numérique

Unité de recherche d'accueil

Unité de Recherche : Inria/IRISA

Equipe : LACODAM

Encadrement

Directeur de thèse : Alexandre Termier **Unité de recherche** : IRISA

Employeur/Établissement d'origine : Université de Rennes I

HDR : • oui • non **Nb de thèses en cours** : 6[1] (4 au début de la thèse)

Mail : alexandre.termier@irisa.fr **Tél.** : +33 2 99 84 71 13

Co-directeur de thèse : Luis Galárraga **Unité de recherche** : IRISA

Employeur/Établissement d'origine : Inria

HDR : • oui • non **Nb de thèses en cours** : 3

Mail : luis.galarraga@inria.fr **Tél.** : +33 2 99 84 71 29

Sujet

Titre : Mine once, Update sometimes : Energy-Efficient Rule Mining on Knowledge Graphs via Provenance

Mots clés : knowledge graphs, provenance, rule mining, energy efficient, incremental

Cotutelle internationale (oui/non) : non **Si oui, préciser l'établissement et le pays de rattachement** :

Descriptif (15-30 lignes) :

Contexte socioéconomique et scientifique ; Les hypothèses et questions posées ; Les grandes étapes de la thèse et démarche ; Approches méthodologiques et techniques envisagées

Multiple AI-assisted tasks rely on the ability to learn patterns from knowledge graphs (KGs). Those patterns can take the form of logical rules such as ***birthCountry(x, y) => nationality(x, y)*** (people are usually citizens of the country they were born in), which are automatically discovered or *mined* from KGs using resource-intensive algorithms. Those algorithms assume that KGs are static. This, however, stands in sharp contrast with reality: modern KGs are continuously updated to be in sync with an ever-changing world where countries change governments, wars break out, and new animal species are discovered. This dynamicity can turn previously learned rules obsolete. If one or several countries abolish birthright citizenship, then the confidence of our example rule may decrease, and so our reliance on it. So far, the traditional way to handle data updates when mining KGs is to rerun the mining algorithm on the latest version of the KG. This can be prohibitively resource-intensive for large and dynamic KGs. We therefore envision a mechanism to mine the rules *incrementally* so that the computational cost of data updates is proportional to the size of the changes -- and not to the size of the data. We call this philosophy *mine once, update sometimes*.

Envisioned Solution and Research Questions

Incremental rule mining is an instance of *update propagation* for query results in databases. This is so because pattern mining relies on executing many queries on the data. For instance, the *support* of our example rule (which determines partly how much we should trust the rule) is the number of results of query **Q(x, y) := *birthCountry(x, y), nationality(x, y)***. If the KG changes, the support of this query may also change. To update the query's solution count without rerunning the query from scratch, the literature in databases has proposed to propagate updates via *how-provenance explanations*. These are expressions that convey a trace of the interactions between the KG records that contribute to a given query solution. These explanations can be used as proofs to verify if a solution is still a solution after some changes in the data. Based on this observation, we propose to carry out rule mining once, store the mined rules, and then propagate subsequent updates from the KG to the different scores computed on the rules – in the spirit of the philosophy *mine once, update sometimes*. This thesis is therefore built upon two main research questions: (a) What is a good representation for the KG, the rules, and the provenance explanations in an incremental mining setting?; (b) how much time and energy consumption can be saved with a mine-once-update-sometimes approach to rule mining on KGs?.

➢ **Ce document est à renvoyer par la direction de l'unité de Recherche aux animateurs du Pole Recherche et à l'école Doctorale pour le 6 février 2023 au plus tard**

1 En automne 2023, Lénaïg Cornanguer et Victor Guyomard soutiendront leurs thèses, le nombre de thèses au démarrage de celle de ce projet sera donc de 4.